The norm of products of free random variables

Vladislav Kargin

Abstract Let X_i denote free identically-distributed random variables. This paper investigates how the norm of products $\Pi_n = X_1 X_2 \cdots X_n$ behaves as n approaches infinity. In addition, for positive X_i it studies the asymptotic behavior of the norm of $Y_n = X_1 \circ X_2 \circ \cdots \circ X_n$, where $\circ$ denotes the symmetric product of two positive operators: $A \circ B =: A^{1/2} B A^{1/2}$. It is proved that if $E X_i = 1$, then $\|Y_n\|$ is between $c_1 \sqrt{n}$ and $c_2 n$ for certain constant c_1 and c_2. For $\|\Pi_n\|$, it is proved that the limit of $n^{-1} \log \|\Pi_n\|$ exists and equals $\log \sqrt{E\left(X_i^* X_i\right)}$. Finally, if π is a cyclic representation of the algebra generated by X_i, and if ξ is a cyclic vector, then $n^{-1} \log \|\pi\left(\Pi_n\right)\xi\| = \log \sqrt{E\left(X_i^* X_i\right)}$ for all n. These results are significantly different from analogous results for commuting random variables.

Mathematics Subject Classification (2000)   46L54 · 15A52

1 Introduction

Suppose $X_1, X_2, \ldots, X_n$ are identically-distributed free random variables. These variables are infinite-dimensional linear operators but the reader may find it convenient to think of them as very large random matrices. The first question we will address in this paper is how the norm of $\Pi_n = X_1 X_2 \cdots X_n$ behaves. If X_i are all positive, then it is natural to look also at the symmetric product operation $\circ$ defined as follows: $X_1 \circ X_2 = X_1^{1/2} X_2 X_1^{1/2}$. The benefit is that unlike the usual operator product, this operation maps the set of positive variables

V. Kargin (✉)
Courant Institute of Mathematical Sciences, 109-20 71st Road, Apt. 4A,
Forest Hills, NY 11375, USA
e-mail: kargin@cims.nyu.edu

to itself. For this operation we can ask how the norm of symmetric products $Y_n = X_1 \circ X_2 \circ \cdots \circ X_n$ behaves.[1]

Products of random matrices and their asymptotic behavior were originally studied by Bellman [1]. One of the decisive steps was made by Furstenberg and Kesten [5], who investigated a matrix-valued stationary stochastic process $X_1, \ldots, X_n, \ldots$, and proved that the limit of $n^{-1} E (\log \|X_1 \cdots X_n\|)$ exists (but might equal $\pm\infty$) and that under certain assumptions $n^{-1} \log \|X_1 \cdots X_n\|$ converges to this limit almost surely. Essentially, the only facts that are used in the proof of this result are the ergodic theorem, the norm inequality $\|X_1 X_2\| \leq \|X_1\| \|X_2\|$ and the fact that the unit sphere is compact in finite-dimensional spaces. It is the lack of compactness of the unit sphere in the infinite-dimensional space that makes generalizations to infinite-dimensional operators non-trivial (see Ruelle [11] for a generalization in the case of compact operators). More work on non-commutative products was done by Furstenberg [4], Oseledec [10], Kingman [8], and others. The results are often called multiplicative ergodic theorems and they find many applications in mathematical physics. For example, see Ruelle [12].

In this paper, we study products of free random variables. These variables are (non-compact) infinite-dimensional operators which can be thought of as a limiting case of large independent random matrices.

Suppose that X_i are free, identically-distributed, self-adjoint, and positive. Suppose also $E (X_i) = 1$. Then we show that the norm of $Y_n = X_1 \circ X_2 \circ \cdots \circ X_n$ grows no faster than a linear function of n. Precisely, we find that

$$\lim_{n \to \infty} \sup n^{-1} \|Y_n\| \leq c_1 \|X_i\|.$$

We are also able to show that if X_i is not concentrated at 1, then

$$\lim_{n \to \infty} \inf n^{-1/2} \|Y_n\| \geq c_2 > 0.$$

For the usual products $\Pi_n = X_1 X_2 \cdots X_n$ we can relax the assumption of self-adjointness. So, suppose that X_i are free and identically-distributed but not necessarily self-adjoint. Also, we do not require that $E (X_i) = 1$. Then we show that

$$\lim_{n \to \infty} n^{-1} \log \|\Pi_n\| = \log \sqrt{E (X_i^* X_i)}. \tag{1}$$

Another way to describe the behavior of Π_n is to look at how the norm of a fixed vector ξ changes when we consecutively apply free operators $X_1, \ldots, X_n$

---

[1] The operation $\circ$ is neither commutative, nor associative. By convention we multiply starting on the right, so, for example, $X_1 \circ X_2 \circ X_3 \circ X_4 = X_1 \circ (X_2 \circ (X_3 \circ X_4))$. However, this convention is not important for the question that we ask. First, it is easy to check that $X_1 \circ X_2$ has the same spectral distribution and therefore the same norm as $X_2 \circ X_1$. Second, if $X_1, X_2,$ and X_3 are free, then the spectral distribution of $(X_1 \circ X_2) \circ X_3$ is the same as the spectral distribution of $X_1 \circ (X_2 \circ X_3)$, and therefore these two products have the same norm. In brief, if X_i are free, then the norm of $X_1 \circ X_2 \circ \cdots \circ X_n$ does not depend on the order in which X_i are multiplied by the operation $\circ$.

to it. More precisely, suppose that the action of the algebra of variables X_i on a Hilbert space H is described by a cyclic representation π and that the vector ξ is cyclic with respect to the expectation E. By definition, this means that $E(X) = \langle \xi, \pi(X)\xi \rangle$ for every operator X from a given algebra. Then we show that

$$n^{-1} \log \| \pi(\Pi_n)\xi \| = \log \sqrt{E(X_i^* X_i)}. \tag{2}$$

Note that we do not need to take the limit, since the equality holds for all n.

The reader may think of cyclic vectors as typical vectors. For example, if the representation π is cyclic and irreducible then cyclic vectors are dense in H. In colloquial terms, (1) says that for large n the product Π_n cannot increase the norm of any given vector ξ by more than $\left[E(X^* X) \right]^{n/2}$. And (2) says that for every cyclic vector ξ this growth rate is achieved.

One more way to capture the intuition of this result is to write

$$\lim_{n \to \infty} n^{-1} \log \| \Pi_n \| = \lim_{n \to \infty} n^{-1} \log \sup_{\|x\|=1} \| \pi(\Pi_n)x \|$$

We have shown that this limit is equal to

$$n^{-1} \log \| \pi(\Pi_n)\xi \|$$

where ξ is a cyclic vector. Thus, for large n the product Π_n acts uniformly in all directions. Its maximal dilation as measured by $\sup_{\|x\|=1} \| \pi(\Pi_n)x \|$ has the same exponential order of magnitude as the dilation in the direction of a typical vector ξ.

It is helpful to compare these results with the case of commutative random variables. Suppose for the moment that X_i are independent commutative random variables with positive values. Then,

$$\lim_{n \to \infty} n^{-1} \log \| X_1 \cdots X_n \| = \log \| X_i \|,$$

where the norm of a random variable is the essential supremum norm (i.e., $\|X\| = \operatorname{ess\,sup}_{\omega \in \Omega} |X(\omega)|$). Indeed, for every $\varepsilon > 0$ the measure of the set $\{\omega : |X_1(\omega) \cdots X_n(\omega)| \geq \|X_1\| \cdots \|X_n\| - \varepsilon\}$ is positive. Therefore $\|X_1 \cdots X_n\| = \|X_1\|^n$. Note that $\log \sqrt{E(X_i^* X_i)} \leq \log \|X_i\|$ and therefore the norm of free products grows more slowly than we would expect from the classical case.

Another interesting comparison is that with results about products of random matrices. Let X_i be i.i.d. random $k \times k$ matrices. Then under suitable conditions, $\lim_{n \to \infty} n^{-1} \log \| X_n \cdots X_1 \|$ exists almost surely. Let us denote this limit as λ. Furstenberg [4] developed a general formula for λ, and Cohen and Newman [3] derived explicit results in the case when entries of X_i have a joint Gaussian distribution. In particular, if all entries of X_i are independent and have the distribution $\mathcal{N}(0, s_k^2)$ then $\lambda = (1/2)\{\log(s_k^2) + \log 2 + \psi(k/2)\}$ where ψ is the digamma function ($\psi(x) = \mathrm{d} \log \Gamma(x)/\mathrm{d}x$). If the size of the matrices grows

$(k \to \infty)$ then $\lambda \sim (1/2) \log \left(k s_k^2\right)$. To compare this with our results, note that if $k s_k^2 \to s^2$, then the sequence of random matrices approximates a free random variable $\widetilde{X}_i$ with the spectral distribution that is uniform inside the circle of radius s. For this free variable, $E\left(\widetilde{X}_i^* \widetilde{X}_i\right) = s^2$, and our theorem shows that $\lim_{n \to \infty} n^{-1} \left\|\widetilde{X}_1 \cdots \widetilde{X}_n\right\| = \log s$. This limit agrees with the result for random matrices. Thus, our result can be seen as a limiting form of results for random matrices.

The results regarding $\|Y_n\|$ are also interesting. We can associate with X_i and Y_n probability measures μ_X and μ_{Y_n}, which are called the spectral probability measures of X_i and Y_n, respectively. Then the measure μ_{Y_n} is determined only by n and the measure μ_X and is called the n-time *free multiplicative convolution* of μ_X with itself:

$$\mu_{Y_n} = \underbrace{\mu_X \boxtimes \cdots \boxtimes \mu_X}_{n \text{ times}}.$$

The norm $\|Y_n\|$ is easy to interpret in terms of the distribution μ_{Y_n}. Indeed, it is the smallest number t such that the support of μ_{Y_n} is inside the interval $[0, t]$. Therefore, the growth in $\|Y_n\|$ measures the growth in the support of the spectral probability measure if the measure is convolved with itself using the operation of the free multiplicative convolution.

In the case of classical multiplicative convolutions of probability measures, the support grows exponentially, so that if μ_X is supported on $[0, L_X]$, then the measure $\mu_{X_1 \cdots X_n}$ is supported on $\left[0, (L_X)^n\right]$. What we have found in the case of free multiplicative convolutions is that if we fix $E X_i = 1$, then the support of the μ_{Y_n} grows no faster than a linear function of n, i.e., the support of μ_{Y_n} is inside the interval $[0, c n L_x]$ with an absolute constant c.

As was pointed out in the literature, a similar phenomenon occurs for sums of free random variables. The support of measures obtained by free additive convolutions grows much more slowly than in the case of classical additive convolutions. This effect was called superconvergence by Bercovici and Voiculescu [2]. Our finding about $\|Y_n\|$ can be considered as a superconvergence for free multiplicative convolutions.

The rest of the paper is organized as follows. Section 2 formulates the results. Section 3 contains the necessary technical background from free probability theory. Sections 4, 5, and 6 prove the results. And Sect. 7 concludes.

2 Results

A *non-commutative probability space* $(\mathcal{A}, E)$ is a unital C^*-algebra $\mathcal{A}$ and a positive linear functional E, such that $E(I) = 1$. We will assume that the functional is tracial, i.e., $E(AB) = E(BA)$ for any two operators A and B from algebra $\mathcal{A}$. The elements of algebra $\mathcal{A}$ are called *random variables* and the functional E is called the *expectation*. The numbers $E\left(X^k\right)$ are called *moments* of the random variable X.

A prototypical example of a non-commutative probability space is a group algebra. That is, for a countable group G we consider the Hilbert space $L^2(G, \nu)$, where ν is a counting measure, and consider the left action of G on $L^2(G, \nu)$: if $f \in L^2(G, \nu)$ and $a, b \in G$, then $[af](b) = f(ab)$. The elements of the group algebra $\mathcal{A}$ are finite sums $\sum_{a \in G} x_a a$ and we can extend by linearity the action of the group G on $L^2(G, \nu)$ to the action of the algebra $\mathcal{A}$ on $L^2(G, \nu)$. We can additionally complete the resulting operator algebra in an appropriate topology. The expectation of an element $\sum x_a a$ is defined as x_e, where e is the identity of the group.

Another important example is the algebra of random matrices. The expectation of an element X in this algebra is defined as $E(X) = \mathcal{E}(N^{-1} \operatorname{tr}(X))$, where $\mathcal{E}$ is the expectation with respect to underlying randomness and N is the dimension of the random matrix. For more details about these examples the reader may consult Hiai and Petz [7].

The concept of freeness substitutes for the concept of independence. Consider sub-algebras $\mathcal{A}_1, \ldots, \mathcal{A}_n$ be given. Let a_i denote elements of these sub-algebras and let $a_i \in \mathcal{A}_{k(i)}$.

Definition 1 *The algebras $\mathcal{A}_1, \ldots, \mathcal{A}_n$ (and their elements) are called* free, *if $E(a_1 \cdots a_m) = 0$ in all cases when the following two conditions are satisfied:*

i) $E(a_i) = 0$ *for all* $i = 1, \ldots, m$, *and*
ii) $k(m) \neq k(1)$ *and* $k(i) \neq k(i+1)$ *for every* $i < m$.

Consider the group algebra for a free group with at least two generators. Then the operators corresponding to generators are free in the sense of the previous definition. For the algebra of large random matrices, Voiculescu proved the asymptotic freeness of two classically independent Gaussian matrices, where asymptotic means that the property in the previous definition is approached as the dimension of matrices $N \to \infty$ (see Voiculescu [14]).

It turns out that many concepts of classical probability theory can be transferred to the case of free random variables. For example, for a self-adjoint variable we can define its distribution function. Indeed, if A is a self-adjoint operator then by the spectral decomposition theorem it can be written as

$$A = \int_{-\infty}^{\infty} \lambda P(\mathrm{d}\lambda),$$

where P is a positive, projector-valued measure, i.e., a mapping that sends sets of the real axis to orthogonal projectors. This allows definition of the *spectral measure* of A, μ_A, which is a measure with the following distribution function:

$$\mathcal{F}_A(t) = E\left(\int_{-\infty}^{t} P(\mathrm{d}\lambda)\right).$$

We can calculate the expectation of any summable function of a self-adjoint variable A by using its spectral measure:

$$Ef(A) = \int_{-\infty}^{\infty} f(\lambda) \, d\mu_A(\lambda).$$

Let $X_1, X_2, \ldots, X_n$ be free identically-distributed positive random variables. Consider $\Pi_n = X_1 X_2 \cdots X_n$ and $Y_n = X_1 \circ X_2 \circ \cdots \circ X_n$ (by convention we multiply on the left, so that, for example, $X_1 \circ X_2 \circ X_3 \circ X_4 = X_1 \circ (X_2 \circ (X_3 \circ X_4)))$. We will see later that these variables have the same moments: $E(\Pi_n)^k = E(Y_n)^k$. As a first step let us record some simple results about the expectation and variance of Y_n and Π_n. We define *variance* of a random variable A as

$$\sigma^2(A) =: E(A^2) - (E(A))^2.$$

Proposition 1 *Suppose that X_i are self-adjoint and $E(X_i) = 1$. Then $E(\Pi_n) = E(Y_n) = 1$ and $\sigma^2(\Pi_n) = \sigma^2(Y_n) = n\sigma^2(X_i)$.*

Note that the linear growth in the variance of $\Pi_n = X_1 \cdots X_n$ is in contrast with the classical case, where only the variance of $\log(X_1 \cdots X_n)$ grows linearly. We will prove this Proposition later when we have more technical tools available. Before that we are going to formulate the main results.

Let $\|A\|$ denote the usual operator norm of operator A.

Theorem 1 *Suppose that $X_1, \ldots, X_n$ are identically-distributed positive self-adjoint free variables. Suppose also that $E(X_i) = 1$. Then*

(1)   *there exists such a constant, c, that $\|Y_n\| \leq c \|X_i\| \, n$;*
      and
(2) $\|Y_n\| \geq \sigma(X_i) \sqrt{n}$.

For the next theorem define

$$\gamma = \sigma\left(\frac{X_i^* X_i}{E(X_i^* X_i)}\right) \geq 0$$

Theorem 2 *Suppose that $X_1, \ldots, X_n$ are free identically-distributed variables* (*not necessarily self-adjoint*). *Then*

(1)   *there exists such a constant, c, that $\|\Pi_n\| \leq c \|X_i\| \sqrt{n} \left[E(X_i^* X_i)\right]^{(n-1)/2}$;*
      and
(2) $\|\Pi_n\| \geq \gamma^{1/2} n^{1/4} \left[E(X_i^* X_i)\right]^{n/2}$.

Corollary 1 *Suppose that $X_1, \ldots, X_n$ are free identically-distributed variables* (*not necessarily self-adjoint*). *Then*

$$\lim_{n \to \infty} n^{-1} \log \|\Pi_n\| = \log \sqrt{E(X_i^* X_i)}$$

Next, suppose that the algebra $\mathcal{A}$ acts on an (infinitely-dimensional) Hilbert space H. In other words, let π be a representation of $\mathcal{A}$. We call representation π *cyclic* if there exists such a vector $\xi \in H$ that $E(X) = \langle \xi, \pi(X)\xi \rangle$ for all operators $X \in \mathcal{A}$. The vectors with this property are also called *cyclic*.

Theorem 3 *Suppose π is a cyclic representation of $\mathcal{A}$, ξ is its cyclic vector, and $X_1, \ldots, X_n$ are free identically-distributed variables from $\mathcal{A}$. Then*

$$n^{-1} \log \|\pi(\Pi_n)\xi\| = \log \sqrt{E(X_i^* X_i)}$$

Corollary 2 *If π and ξ are cyclic then*

$$\log \|\Pi_n\| \sim \log \|\pi(\Pi_n)\xi\| \sim n \log \|\pi(X_1)\xi\|$$

as $n \to \infty$.

3 Preliminaries

The *Cauchy transform* of a bounded random variable A is defined as follows:

$$G_A(z) = E\left(\frac{1}{z-A}\right) = \frac{1}{z} + \sum_{k=1}^{\infty} \frac{E(A^k)}{z^{k+1}}.$$

This power series is convergent for $|z| > \|A\|$. Let us also define the ψ*-function* of A:

$$\psi_A(z) = E\left(\frac{1}{1-zA}\right) - 1 = \sum_{k=1}^{\infty} E(A^k) z^k.$$

The ψ-function is convergent for $|z| \leq \|A\|^{-1}$ and it is related to the Cauchy transform by the following equality:

$$G_n(z) = z^{-1}\left[\psi_n\left(z^{-1}\right) + 1\right].$$

If A is bounded and $E(A) \neq 0$, then for z in a sufficiently small neighborhood of 0, the inverse of $\psi_A(z)$ is defined, which we denote as $\psi_A^{-1}(z)$. Then the *S-transform* is defined as

$$S_A(z) = \left(1 + \frac{1}{z}\right)\psi_A^{-1}(z) \tag{3}$$

Let us write out several first terms in the power expansions for $\psi(z), \psi^{-1}(z)$, and $S(z)$. Suppose for simplicity that $E(A) = 1$ and let $E(A^k) = m_k$. Then,

$$\psi(z) = z + m_2 z^2 + m_3 z^3 + \cdots,$$
$$\psi^{-1}(z) = z - m_2 z^2 - \left(m_3 - 2m_2^2\right) z^3 + \cdots,$$
$$S(z) = 1 + (1 - m_2)z + \left(2m_2^2 - m_2 - m_3\right) z^2 + \cdots$$

The main theorem regarding the multiplication of free random variables was proved by Voiculescu [13]. Later the proof was significantly simplified by Haagerup [6].

Theorem 4 (Voiculescu) *Suppose X and Y are bounded free random variables. Suppose also that $E(X) \neq 0$ and $E(Y) \neq 0$. Then*

$$S_{XY}(z) = S_X(z) S_Y(z).$$

In particular, this theorem implies that $S_{\Pi_n} = S_{Y_n} = (S_X)^n$, where S_X denotes the S-transform of any of X_i. Now it is easy to prove Proposition 1. Indeed, let us denote S_{Π_n} as S_n. Then, using the power expansions we can write:

$$S_n(z) = 1 + \left(1 - m_2^{(n)}\right) z + \cdots$$
$$= (S_X)^n = 1 + n(1 - m_2) z + \cdots,$$

where $m_2^{(n)} =: E(\Pi_n)^2$ and $m_2 =: E(X_i)^2$. Then, using power expansion in (3), we conclude that $E(\Pi_n) = 1$. Next, by definition, $\sigma^2(X_i) = m_2 - 1$ and $\sigma^2(\Pi_n) = m_2^{(n)} - 1$. Therefore, we can conclude that $\sigma^2(\Pi_n) = n\sigma^2(X)$.

4 Proof of theorem 1

Throughout this section we assume that X_i are self-adjoint, $E(X_i) = 1$, and the support of the spectral distribution of X_i belongs to $[0, L]$.

Let us first go in a simpler direction and derive a lower bound on $\|Y_n\|$. That is, we are going to prove claim (2) of the theorem. From Proposition 1, we know that $E(Y_n) = 1$ and $\sigma^2(Y_n) = n\sigma^2(X_i)$. It is clear that for every positive random variable A, it is true that $E(A^2) \leq \|A\|^2$ and therefore $\|A\| \geq \sqrt{\sigma^2(A) + [E(A)]^2}$. Applying this to Y_n, we get $\|Y_n\| \geq \sqrt{n\sigma^2 + 1}$. In particular, $\|Y_n\| > \sigma\sqrt{n}$, so (2) is proved.

Now let us prove claim (1). By Theorem 4, $S_n(z) = (S_X(z))^n$. The idea of the proof is to investigate how $|S_X(z)|^n$ behaves for small z. It turns out that if z is of the order of n^{-1}, then $|S_X(z)|^n > c$ where c is a constant that does not depend on n. We will show that this fact implies that $\psi_n(z)$ (i.e., the ψ-function for Y_n) has the convergent power series in the area $|z| < (cn)^{-1}$ and that therefore the Cauchy transform of Y_n has the convergent power series in $|z| > cn$.

This fact and the Perron-Stieltjes inversion formula imply that the support of the distribution of Y_n is inside $[-cn, cn]$.

In the proof we need the result about functional inversions formulated below. By a function holomorphic in a domain, D, we mean a function which is bounded and differentiable in D.

Lemma 1 (Lagrange's inversion formula) *Suppose f is a function of a complex variable, which is holomorphic in a neighborhood of $z_0 = 0$ and has the Taylor expansion*

$$f(z) = z + \sum_{k=2}^{\infty} a_k z^k + \cdots,$$

converging for all sufficiently small z. Then the functional inverse of $f(z)$ is well defined in a neighborhood of 0 and the Taylor series of the inverse is given by the following formula:

$$f^{-1}(u) = u + \sum_{k=2}^{\infty} \left[\frac{1}{2\pi i k} \oint_{\gamma} \frac{dz}{f(z)^k} \right] u^k,$$

where γ is a circle around 0, in which f has only one zero.

For the proof see Theorems II.3.2 and II.3.3 in Markushevich [9], or Section 7.32 in Whittaker and Watson [15].

Lemma 2 $E\left(X^k\right) \leq L^{k-1}$.

Proof

$$E\left(X^k\right) = \int_0^L \lambda^k d\mu_X(\lambda) \leq L^{k-1} \int_0^L \lambda d\mu_X(\lambda) = L^{k-1},$$

where $d\mu_X$ denotes the spectral distribution of the variable X. $\qquad\square$

Lemma 3 *The function $\psi_X(z)$ is has only one zero in $|z| \leq (4L)^{-1}$, and if $|z| = (4L)^{-1}$, then $\left|\psi_X(z)\right| \geq (6L)^{-1}$.*

Proof If $|z| \leq (4L)^{-1}$ then

$$\left|\psi_X(z) - z\right| \leq |z| \sum_{k=2}^{\infty} E\left(X^k\right) |z|^{k-1}$$

$$\leq |z| \sum_{k=1}^{\infty} \frac{1}{4^k} = \frac{|z|}{3}.$$

Therefore, by Rouché's theorem, $\psi_X(z)$ has only one zero in this area.

If $|z| = (4L)^{-1}$, then

$$\left|\psi_X(z)\right| \geq |z| - \sum_{k=2}^{\infty} E\left(X^k\right) |z|^k$$

$$\geq |z| \left(1 - \sum_{k=1}^{\infty} \frac{1}{4^k}\right)$$

$$= \frac{1}{4L}\left(1 - \frac{1}{3}\right) = \frac{1}{6L}.$$

$\square$

By Lemma 1, we can expand the functional inverse of $\psi_X(z)$ as follows:

$$\psi_X^{-1}(u) = u + \sum_{k=2}^{\infty} c_k u^k,$$

where

$$c_k = \frac{1}{2\pi i k} \int_\gamma \frac{dz}{\left[\psi_X(z)\right]^k}$$

Lemma 4 *If $|u| \leq (72Ln)^{-1}$, then*

$$\left|\frac{\psi_X^{-1}(u)}{u} - 1\right| \leq \frac{1}{7n}.$$

Proof Using the previous lemma we can estimate c_k:

$$c_k \leq \frac{1}{k}\frac{1}{4L}(6L)^k \leq \frac{3}{2}(6L)^{k-1}.$$

Then

$$\left|\frac{\psi_X^{-1}(u)}{u} - 1\right| = \left|\sum_{k=2}^{\infty} c_k u^{k-1}\right|$$

$$\leq \frac{3}{2}\sum_{k=1}^{\infty}\left(\frac{1}{12n}\right)^k = \frac{3}{2}\frac{1}{12n-1}$$

$$= \frac{3}{2}\frac{12n}{12n-1}\frac{1}{12n} \leq \frac{1}{7n},$$

provided that $|u| \leq (72Ln)^{-1}$.

$\square$

Lemma 5 *If $|u| \leq (72Ln)^{-1}$, then*

$$|1 - S_X(u)| \leq \frac{1}{6n}.$$

Proof Recall that $S_X(u) = (1+u)\,\psi_X^{-1}(u)/u$. Then we can write:

$$
\begin{aligned}
|1 - S_X(u)| &= \left| u + (1+u)\left(\frac{\psi_X^{-1}(u)}{u} - 1\right)\right| \\
&\leq |u| + |1+u|\left|\frac{\psi_X^{-1}(u)}{u} - 1\right|.
\end{aligned}
$$

Then the previous lemma implies that for $|u| \leq (72Ln)^{-1}$ and $n \geq 2$, we have the estimate:

$$|1 - S_X(u)| \leq \frac{1}{72Ln} + \left|1 + \frac{1}{72Ln}\right|\frac{1}{7n}.$$

Note that $L \geq 1$ because $EX = 1$. Therefore,

$$|1 - S_X(u)| \leq \frac{1}{72n} + \frac{73}{72}\frac{1}{7n} \leq \frac{1}{6n}.$$

$\square$

Lemma 6 *For all positive integer n if $|u| \leq (72Ln)^{-1}$, then*

$$e^{1/6} \geq |S_X(u)|^n \geq e^{-1/3}.$$

Proof Let us first prove the upper bound on $|S_X(u)|^n$. The previous lemma implies that

$$|S_X(u)|^n \leq \left(1 + \frac{1}{6n}\right)^n \leq e^{1/6}.$$

Now let us prove the lower bound. The previous lemma implies that

$$|S_X(u)|^n \geq \left(1 - \frac{1}{6n}\right)^n.$$

In an equivalent form,

$$n\log|S_X(u)| \geq n\log\left(1 - \frac{1}{6n}\right). \tag{4}$$

Recall the following elementary inequality: If $x \in [0, 1 - e^{-1}]$, then

$$\log (1 - x) \geq -2x.$$

Let $x = 1/(6n)$. Then

$$\log \left(1 - \frac{1}{6n}\right) \geq -\frac{1}{3n}.$$

Substituting this in (4), we get

$$n \log |S_X(u)| \geq -\frac{1}{3},$$

or

$$|S_X(u)|^n \geq e^{-1/3}.$$

□

By Theorem 4, $S_n(u) =: [S_X(u)]^n$ is the S-transform of the variable Y_n. The corresponding inverse ψ-function is $\psi_n^{-1}(u) = uS_n(u)/(1+u)$.

First, we estimate $S_n(u) - 1$.

Lemma 7 *If $|u| \leq (72Ln)^{-1}$, then*

$$|S_n(u) - 1| \leq \frac{1}{5}.$$

Proof Write

$$\left|S_X(u)^n - 1\right| \leq |S_X(u) - 1| \left(|S_X(u)|^{n-1} + |S_X(u)|^{n-2} + \cdots + 1\right)$$
$$\leq \frac{1}{6n} e^{1/6} n \leq \frac{1}{5}.$$

□

Lemma 8 *The function $\psi_n^{-1}(u)$ has only one zero in $|u| = (72Ln)^{-1}$ and if $|u| = (72Ln)^{-1}$, then*

$$\left|\psi_n^{-1}(u)\right| \geq \frac{1}{102Ln}.$$

Proof Recall that by definition in (3), $\psi_n^{-1}(u) = uS_n(u)/(1+u)$. Therefore,

$$\left|\psi_n^{-1}(u) - u\right| = |u| \left|\frac{S_n(u) - (1+u)}{1+u}\right|$$

and by Lemma 7 we have the following estimate:

$$\left| \frac{S_n(u) - (1+u)}{1+u} \right| \leq \frac{1}{1-|u|} |S_n(u) - 1| + \frac{|u|}{1-|u|}$$
$$\leq \frac{72}{71} \frac{1}{5} + \frac{1}{71} \leq \frac{1}{4}.$$

Therefore, by Rouché's theorem, $\psi_n^{-1}(u)$ has only one zero in $|u| \leq (72Ln)^{-1}$.
Next, note that $\psi_n^{-1}(u) = uS_n(u) / (1+u)$ and if $|u| = (72Ln)^{-1}$, then

$$\left| \frac{u}{1+u} \right| \geq \frac{1}{72Ln} / \left(1 + \frac{1}{72Ln} \right) \geq \frac{1}{73Ln}.$$

Using Lemma 6, we get:

$$\left| \psi_n^{-1}(u) \right| \geq \frac{1}{73Ln} e^{-1/3} \geq \frac{1}{102Ln}.$$

□

Now we again apply Lemma 1 and obtain the following formula:

$$\psi_n(z) = z + \sum_{k=2}^{\infty} \left[\frac{1}{2\pi i k} \oint_\gamma \frac{du}{\left[\psi_n^{-1}(u) \right]^k} \right] z^k, \tag{5}$$

where we can take the circle $|u| = (72Ln)^{-1}$ as γ.

Lemma 9 *The radius of convergence of series* (5) *is at least* $(102Ln)^{-1}$.

Proof By the previous lemma, the coefficient before z^k can be estimated as follows:

$$|c_k| \leq \frac{1}{k} \frac{1}{72Ln} (102Ln)^k.$$

This implies that series (5) converges at least for $|z| \leq (102Ln)^{-1}$.   □

Lemma 10 *The support of the spectral distribution of* $Y_n = X_1 \circ X_2 \circ \cdots \circ X_n$
belongs to the interval $[-102Ln, 102Ln]$.

Proof The variable Y_n is self-adjoint and has a well-defined spectral measure, $\mu_n(dx)$, supported on the real axis. We can infer the Cauchy transform of this measure from $\psi_n(z)$:

$$G_n(z) = z^{-1} \left[\psi_n\left(z^{-1} \right) + 1 \right].$$

Using Lemma 9, we can conclude that the power series for $G_n(z)$ around $z = \infty$ converges in the area $|z| > 102Ln$. The coefficients of this series are real. Therefore, using the Perron-Stieltjes formula we conclude that $\mu_n(dx)$ is zero outside of the interval $[-102Ln, 102Ln]$. □

Lemma 10 implies the statement of Theorem 1.

5 Proof of theorem 2

The norm of the operator Π_n coincides with the square root of the norm of the operator $\Pi_n^* \Pi_n$. Therefore, all we need to do is to estimate the norm of the self-adjoint operator $\Pi_n^* \Pi_n$.

Lemma 11 *For every bounded operator $X \in \mathcal{A}$, products X^*X and XX^* have the same spectral distribution.*

Proof Since E is tracial, $E(X^*X)^k = E(XX^*)^k$. Therefore, X^*X and XX^* have the same sequence of moments and, therefore, the same distribution. □

If two variables A and B have the same sequence of moments, we say that they are *equivalent* and write $A \sim B$. In particular, two self-adjoint bounded variables have the same spectral distribution if and only if they are equivalent.

Lemma 12 *Let $A, B,$ and C be three bounded operators from a non-commutative probability space $\mathcal{A}$. If $A \sim B$, A is free from C, and B is free from C, then $A + C \sim B + C$, $AC \sim BC$, and $CA \sim CB$.*

Proof Since A and C are free, the moments of $A + C$ can be computed from the moments of A and C. The computation is exactly the same for $B + C$, since B and C are also free. In addition, we know that A and B have the same moments. Consequently, $A + C$ has the same moments as $B + C$, i.e., $A + C \sim B + C$. The other equivalences are obtained similarly. □

Lemma 13 *If $A \sim B$, then $S_A(z) = S_B(z)$. In words, if two variables are equivalent, then they have the same S-transform.*

Proof From the definition of the ψ-function, it is clear that if $A \sim B$, then $\psi_A(z) = \psi_B(z)$. This implies that $\psi_A^{-1}(z) = \psi_B^{-1}(z)$ and therefore $S_A(z) = S_B(z)$. □

For example, $S_{X_i^* X_i}(z)$ does not depend on i and we will denote this function as $S_{X^*X}(z)$.

Lemma 14 *If $X_1, \ldots, X_n$ are free, then*

$$\Pi_n^* \Pi_n \sim X_n^* X_n \cdots X_1^* X_1$$

and if $X_1, \ldots, X_n$ are in addition identically distributed, then

$$S_{\Pi_n^* \Pi_n} = S_{\Pi_n \Pi_n^*} = (S_{X^*X})^n$$

Proof We will use induction. For $n = 1$, we have $\Pi_1^* \Pi_1 = X_1^* X_1$. Therefore $S_{\Pi_1^* \Pi_1} = S_{X^* X}$. Suppose that the statement is proved for $n - 1$. Then

$$\Pi_n^* \Pi_n = X_n^* \cdots X_1^* X_1 \cdots X_n$$
$$\sim X_n X_n^* X_{n-1}^* \cdots X_1^* X_1 \cdots X_{n-1},$$

where the equivalence holds because E is tracial and it is easy to check that the products have the same moments. Therefore,

$$\Pi_n^* \Pi_n \sim \left(X_n X_n^* \right) \Pi_{n-1}^* \Pi_{n-1}$$
$$\sim \left(X_n^* X_n \right) \Pi_{n-1}^* \Pi_{n-1}$$

by Lemmas 11 and 12. Then the inductive hypothesis implies that

$$\Pi_n^* \Pi_n \sim X_n^* X_n \cdots X_1^* X_1.$$

Using Lemma 13 and Theorem 4, we write:

$$S_{\Pi_n^* \Pi_n} = (S_{X^* X})^n.$$

Since $\Pi_n^* \Pi_n \sim \Pi_n \Pi_n^*$, therefore, $S_{\Pi_n^* \Pi_n} = S_{\Pi_n \Pi_n^*} = (S_{X^* X})^n.$ □

We have managed to represent $S_{\Pi_n^* \Pi_n}$ as $(S_{X^* X})^n$ and therefore all the arguments of the previous section are applicable, except that we are interested in $(S_{X^* X})^n$ rather than in $(S_X)^n$. In particular, we can conclude that the following lemma holds:

Lemma 15 *Define*

$$\gamma = \sigma \left(\frac{X_i^* X_i}{E \left(X_i^* X_i \right)} \right).$$

Then

(1) $\left\| \Pi_n^* \Pi_n \right\| \leq 102 \left\| X_i \right\|^2 n E \left(X_i^* X_i \right)^{n-1}$, *and*

(2) $\left\| \Pi_n^* \Pi_n \right\| \geq \gamma \sqrt{n} E \left(X_i^* X_i \right)^n.$

Proof Let us introduce variables $R_i = s^{-1} X_i$, where $s^2 = E (X^* X)$. Then $\left\| R_i^* R_i \right\| = (\left\| X_i \right\| / s)^2$ and $E \left(R_i^* R_i \right) = 1$. Let $\tilde{\Pi}_n = R_1 \cdots R_n$. Then $\Pi_n^* \Pi_n = s^{2n} \tilde{\Pi}_n^* \tilde{\Pi}_n$ and the S-transform of $\tilde{\Pi}_n^* \tilde{\Pi}_n$ is $(S_{R^* R})^n$.

Note that $\tilde{\Pi}_n^* \tilde{\Pi}_n$ has the same S-transform and therefore the same spectral distribution as $\left(R_1^* R_1 \right) \circ \cdots \circ \left(R_n^* R_n \right)$. Using Theorem 1, we conclude that

$\left\| \widetilde{\Pi}_n^* \widetilde{\Pi}_n \right\| \leq 102 \left(\|X_i\| / s \right)^2 n$. It follows that $\left\| \Pi_n^* \Pi_n \right\| \leq 102 \|X_i\|^2 s^{2n-2} n$. In addition, Theorem 1 implies that

$$\left\| \widetilde{\Pi}_n^* \widetilde{\Pi}_n \right\| > \sqrt{n} \sigma \left(R_i^* R_i \right)$$
$$= \gamma \sqrt{n}$$

Consequently,

$$\left\| \Pi_n^* \Pi_n \right\| \geq \gamma \sqrt{n} s^{2n}.$$

$\square$

From Lemma 15 we conclude that

$$\|\Pi_n\| \leq 11 \|X_i\| \sqrt{n} \left[E \left(X_i^* X_i \right) \right]^{(n-1)/2},$$

and

$$\|\Pi_n\| \geq \gamma^{1/2} n^{1/4} \left[E \left(X_i^* X_i \right) \right]^{n/2}$$

This completes the proof of Theorem 2.

6 Proof of theorem 3

By definition of the cyclic vector, we have:

$$\|\pi \left(\Pi_n \right) \xi \|^2 = \langle \pi \left(\Pi_n \right) \xi, \pi \left(\Pi_n \right) \xi \rangle$$
$$= \langle \xi, \pi \left(\Pi_n^* \Pi_n \right) \xi \rangle$$
$$= E \left(\Pi_n^* \Pi_n \right).$$

Using Lemma 14, we continue this as follows:

$$E \left(\Pi_n^* \Pi_n \right) = E \left(X_n^* X_n \cdots X_1^* X_1 \right)$$
$$= \left[E \left(X^* X \right) \right]^n.$$

Consequently,

$$n^{-1} \log \|\Pi_n \xi \| = \frac{1}{2} \log E \left(X^* X \right).$$

$\square$

7 Concluding remarks

We have investigated how the norms of $\Pi_n = X_1 \cdots X_n$ and $Y_n = X_1 \circ \cdots \circ X_n$ grow as $n \to \infty$. For $\|\Pi_n\|$, we have shown that $\lim_{n \to \infty} n^{-1} \log \|\Pi_n\|$ exists and equals $\log \sqrt{E\left(X_i^* X_i\right)}$. For $\|Y_n\|$, we have proved that the growth rate of $\|Y_n\|$ is somewhere between $\sqrt{n}$ and n. There remains the question of whether $\lim_{n \to \infty} n^{-s} \|Y_n\|$ exists for some s.

Another interesting question, which is not resolved in this paper, is how the spectral radius of Π_n grows. Indeed, for Y_n, the norm coincides with the spectral radius. But for Π_n, the norm and the spectral radius are different because Π_n is not self-adjoint.

References

1. Bellman, R.: Limit theorems for non-commutative operations I. Duke Math. J. **21**, 491–500 (1954)
2. Bercovici, H., Voiculescu, D.: Superconvergence to the central limit and failure of the Cramer theorem for free random variables. Prob. Theory Related Fields **102**, 215–222 (1995)
3. Cohen, J.E., Newman, C.M.: The stability of large random matrices and their products. Ann. Prob. **12**, 283–310 (1984)
4. Furstenberg, H.: Noncommuting random products. Trans. Am. Math. Soc. **108**, 377–428 (1963)
5. Furstenberg, H., Kesten, H.: Products of random matrices. Ann. Math. Statist. **31**, 457–469 (1960)
6. Haagerup, U.: On Voiculescus R- and S-transforms for free non-commuting random variables. In: Voiculescu D.-V. (ed.) Free Probability Theory, vol. 12 of Fields Institute Communications, pp. 127–148. American Mathematical Society (1997)
7. Hiai, F., Petz, D.: The Semicircle Law, Free Random Variables And Entropy, 1st edn., vol. 77 of Mathematical Surveys and Monographs. American Mathematical Society (2000)
8. Kingman J., F.C.: Subadditive ergodic theory. Ann. Prob. **1**, 883–899 (1973)
9. Markushevich, A.I.: Theory of Functions of a Complex Variable, 2nd edn. Chelsea Publishing Company, New York (1979)
10. Oseledec, V.I.: A multiplicative ergodic theorem. Ljapunov characteristic numbers for dynamical systems. Trans. Moscow Math. Soc. **19**, 197–231 (1968)
11. Ruelle, D.: Characterisic exponents and invariant manifolds in Hilbert space. Ann. Math. **115**, 243–290 (1982)
12. Ruelle, D.: Characterisic exponents for a viscous fluid subjected to time dependent forces. Commun. Math. Phys. **93**, 285–300 (1984)
13. Voiculescu, D.: Multiplication of certain non-commuting random variables. J. Operator Theory **18**, 223–235 (1987)
14. Voiculescu, D.: Limit laws for random matrices and free products. Invent. Math. **104**, 201–220 (1991)
15. Whittaker, E.T., Watson, G.N.: A Course of Modern Analysis, 4 edn. Cambridge University Press, Cambridge (1927)